Benford's Law - Using Logarithms to Detect Fraud

Darrell Minor, Columbus State Community College

In 1881, Simon Newcomb made the simple observation that the beginning pages of books were more worn than the later pages of those books. From that routine observation, Newcomb and others developed a mathematical principle involving logarithms that can be observed in a wide variety of data, from birth and death rates, to lengths of rivers, to financial transactions. In this article, the author provides an example of how this principle can be used to detect fraud in a company's accounts payable department. Suggestions for classroom activities are provided for additional exploration.

A Fictitious, But Not Fanciful, Example

Darius is the senior accountant at TrueTech-Ohio, a large technology firm that manufactures laboratory equipment for the biomedical industry. It is his responsibility to conduct an internal audit of the company's financial reports every year. As he was reviewing the 2010 reports, he became suspicious that someone in the accounts payable department may be embezzling money from the company, perhaps by writing checks on the company account and then cashing those checks himself or herself.

Newcomb discovered the property upon observing that the beginning pages of books containing tables of logarithms were much more worn than the later pages.

There are only two employees in the accounts payable department who have check-writing authority, Brett and Cynthia. They are scheduled to work on different days - Brett works every Monday and Tuesday; Cynthia works every Thursday and Friday. They alternate working on Wednesdays. Darius decided to sample the checks written during the first two weeks of 2011 in order to analyze them closely for any indication of fraud. He began by reviewing the information in the company's check register shown in Table 1.

At first glance, Darius could find nothing in the check register to cause him concern. Neither Brett nor Cynthia were writing an excessive number of checks (in fact, they had each written 40 checks during the time frame studied), nor were either one of them writing an excessive number of checks for high amounts (they had each written seven checks for over $1,000). Additional analysis that Darius performed turned up nothing unusual - no missing checks, no unusual amounts, etc.

Then Darius remembered reading of a rather obscure property that applies to much real-world data, and that involves logarithms. The property is called "Benford's Law", named after physicist Frank Benford who, in 1938, rediscovered and generalized the law that had previously been formulated by Simon Newcomb in 1881. Newcomb discovered the property upon observing that the beginning pages of books containing tables of logarithms were much more worn than the later pages. This led him to formulate the proposal that in many sets of numerical data, more numbers will tend to begin with a "1" than with any other digit. (Newcomb also questioned, presumably tongue-in-cheek, whether we would observe that books containing tables of anti-logarithms, i.e., exponentials, have the later pages more worn than the earlier pages - Newcomb, 1881).

Table 1 Check register

Date	Check	Amount	Signed by	Date	Check	Amount	Signed by
1/4	2301	39.40	Brett	1/10	2341	15.00	Brett
1/4	2302	94.21	Brett	1/10	2342	36.85	Brett
1/4	2303	108.34	Brett	1/10	2343	68.33	Brett
1/4	2304	276.95	Brett	1/10	2344	185.43	Brett
1/4	2305	491.17	Brett	1/11	2345	578.40	Brett
1/5	2306	14.62	Brett	1/11	2346	761.84	Brett
1/5	2307	198.50	Brett	1/11	2347	805.24	Brett
1/5	2308	217.48	Brett	1/11	2348	1207.11	Brett
1/5	2309	228.35	Brett	1/11	2349	2053.76	Brett
1/5	2310	306.51	Brett	1/11	2350	43.04	Brett
1/5	2311	640.30	Brett	1/11	2351	47.35	Brett
1/5	2312	946.08	Brett	1/11	2352	117.95	Brett
1/5	2313	2115.42	Brett	1/11	2353	847.10	Brett
1/5	2314	3024.66	Brett	1/11	2354	76.95	Brett
1/6	2315	37.25	Cynthia	1/11	2355	284.27	Brett
1/6	2316	56.18	Cynthia	1/11	2356	1820.81	Brett
1/6	2317	62.71	Cynthia	1/12	2357	15.00	Cynthia
1/6	2318	77.40	Cynthia	1/12	2358	34.74	Cynthia
1/6	2319	374.93	Cynthia	1/12	2359	75.34	Cynthia
1/6	2320	455.25	Cynthia	1/12	2360	80.75	Cynthia
1/6	2321	586.23	Cynthia	1/12	2361	504.30	Cynthia
1/6	2322	691.35	Cynthia	1/12	2362	742.01	Cynthia
1/6	2323	903.45	Cynthia	1/12	2363	820.22	Cynthia
1/6	2324	924.07	Cynthia	1/13	2364	895.45	Cynthia
1/6	2325	4115.75	Cynthia	1/13	2365	5435.86	Cynthia
1/7	2326	24.75	Cynthia	1/13	2366	49.10	Cynthia
1/7	2327	86.49	Cynthia	1/13	2367	91.45	Cynthia
1/7	2328	221.99	Cynthia	1/13	2368	164.05	Cynthia
1/7	2329	430.68	Cynthia	1/13	2369	241.11	Cynthia
1/7	2330	2730.19	Cynthia	1/13	2370	1100.43	Cynthia
1/10	2331	19.63	Brett	1/13	2371	1655.36	Cynthia
1/10	2332	25.00	Brett	1/13	2372	6882.41	Cynthia
1/10	2333	59.42	Brett	1/13	2373	28.26	Cynthia
1/10	2334	148.82	Brett	1/13	2374	64.38	Cynthia
1/10	2335	173.10	Brett	1/13	2375	90.53	Cynthia
1/10	2336	374.92	Brett	1/14	2376	318.45	Cynthia
1/10	2337	485.34	Brett	1/14	2377	627.05	Cynthia
1/10	2338	607.21	Brett	1/14	2378	713.55	Cynthia
1/10	2339	1090.86	Brett	1/14	2379	955.23	Cynthia
1/10	2340	5240.46	Brett	1/14	2380	4007.35	Cynthia

This led him to formulate the proposal that in many sets of numerical data, more numbers will tend to begin with a "1" than with any other digit.

In 1938, Benford confirmed Newcomb's observation by analyzing large numbers of data sets, ranging from population data and death rates to newspaper readership and areas of bodies of water (Benford, 1938). He found that the data he analyzed closely followed the general pattern that the smaller the digit, the more frequently it occurred as the first digit in each of his wide variety of data sets.

Benford's Law is also referred to as the "first-digit law," since it involves the distribution frequency of the lead (i.e., first) digits from a data set. Specifically, Benford's Law states that the probability the digit d occurs as the first digit in a set of data is $\log(1 + 1/d)$. Thus, the probability that the lead digit is 1 is $\log(2) \approx 0.30103$; the probability that the lead digit is 2 is $\log(1 + 1/2) = \log(1.5) \approx 0.17609$; etc. Table 2 shows the probability that the lead digit is d for each d value between 1 and 9.

Table 2 Value of $\log(1+1/d)$, d = 1, 2, 3, . . . , 9

d	log(1+1/d)	d	log(1+1/d)	d	log(1+1/d)
1	0.30103	4	0.09691	7	0.05799
2	0.17609	5	0.07918	8	0.05115
3	0.12494	6	0.06695	9	0.04576

In 1938, Benford confirmed Newcomb's observation by analyzing large numbers of data sets, ranging from population data and death rates to newspaper readership and areas of bodies of water (Benford, 1938).

Darius decided to separate the entries from the check register by employee, and then re-order those entries in increasing order of first digits (e.g., 1090.86 would come before 24.75, since "1" comes before "2"). In cases in which the first digits are the same, the entries are ordered in increasing order of second digits (e.g., 3024.66 comes before 37.25); in cases in which the first and the second digit are the same, the entries are ordered in increasing order of third digits; etc. Place values are irrelevant in the ordering of the numbers (so decimal points are essentially ignored).

After compiling the data in this manner, Darius constructed Table 3. Under the bright light of Benford's Law, this data is much more revealing than the data in Table 1. According to Benford's Law, we would expect about 30.1% of the checks written to begin with the digit 1. Since each employee wrote 40 checks, it is reasonable to expect them to have written about 12 checks each for amounts that begin with the digit 1. Brett wrote 12 checks for amounts that begin with the digit 1, while Cynthia only wrote 4 checks for these amounts. Similarly, Benford's Law informs us that we should expect about 17.6% of the checks written to begin with the digit 2; i.e., about 7 checks each. Brett indeed wrote 7 checks that begin with the digit 2, while Cynthia wrote 5. Table 4 shows a summary of the expected number of checks that begin with each digit that we would expect each employee to have written, and the actual number of checks each wrote.

In this hypothetical example, Brett wrote the expected number of checks for each digit. Although that may not be the case in scenarios involving real-world data, the numbers should be reasonably close to the expected values, if not exact, assuming no fraud is occurring. Indeed, Brett's checks demonstrated the expected tendency for the smaller digits to occur more frequently as the initial digit.

Cynthia, on the other hand, had a more evenly-distributed number of checks written for each initial digit, either 4 or 5 in each instance. Thus, we do not observe the expected tendency for the smaller digits to occur more frequently as the initial digit, and have good reason to suspect that Cynthia may be engaging in fraudulent behavior. Of course, Benford's Law alone is not sufficient to draw a final conclusion.

Table 3 Check register ordered by employee and beginning digit

Date	Check	Amount	Signed by	Date	Check	Amount	Signed by
1/4	2303	108.34	Brett	1/13	2370	1100.43	Cynthia
1/10	2339	1090.86	Brett	1/12	2357	15.00	Cynthia
1/11	2352	117.95	Brett	1/13	2368	164.05	Cynthia
1/11	2348	1207.11	Brett	1/13	2371	1655.36	Cynthia
1/5	2306	14.62	Brett	1/7	2328	221.99	Cynthia
1/10	2334	148.82	Brett	1/13	2369	241.11	Cynthia
1/10	2341	15.00	Brett	1/7	2326	24.75	Cynthia
1/10	2335	173.10	Brett	1/7	2330	2730.19	Cynthia
1/11	2356	1820.81	Brett	1/13	2373	28.26	Cynthia
1/10	2344	185.43	Brett	1/14	2376	318.45	Cynthia
1/10	2331	19.63	Brett	1/12	2358	34.74	Cynthia
1/5	2307	198.50	Brett	1/6	2315	37.25	Cynthia
1/11	2349	2053.76	Brett	1/6	2319	374.93	Cynthia
1/5	2313	2115.42	Brett	1/14	2380	4007.35	Cynthia
1/5	2308	217.48	Brett	1/6	2325	4115.75	Cynthia
1/5	2309	228.35	Brett	1/7	2329	430.68	Cynthia
1/10	2332	25.00	Brett	1/6	2320	455.25	Cynthia
1/4	2304	276.95	Brett	1/13	2366	49.10	Cynthia
1/11	2355	284.27	Brett	1/12	2361	504.30	Cynthia
1/5	2314	3024.66	Brett	1/13	2365	5435.86	Cynthia
1/5	2310	306.51	Brett	1/6	2316	56.18	Cynthia
1/10	2342	36.85	Brett	1/6	2321	586.23	Cynthia
1/10	2336	374.92	Brett	1/14	2377	627.05	Cynthia
1/4	2301	39.40	Brett	1/6	2317	62.71	Cynthia
1/11	2350	43.04	Brett	1/13	2374	64.38	Cynthia
1/11	2351	47.35	Brett	1/13	2372	6882.41	Cynthia
1/10	2337	485.34	Brett	1/6	2322	691.35	Cynthia
1/4	2305	491.17	Brett	1/14	2378	713.55	Cynthia
1/10	2340	5240.46	Brett	1/12	2362	742.01	Cynthia
1/11	2345	578.40	Brett	1/12	2359	75.34	Cynthia
1/10	2333	59.42	Brett	1/6	2318	77.40	Cynthia
1/10	2338	607.21	Brett	1/12	2360	80.75	Cynthia
1/5	2311	640.30	Brett	1/12	2363	820.22	Cynthia
1/10	2343	68.33	Brett	1/7	2327	86.49	Cynthia
1/11	2346	761.84	Brett	1/13	2364	895.45	Cynthia
1/11	2354	76.95	Brett	1/6	2323	903.45	Cynthia
1/11	2347	805.24	Brett	1/13	2375	90.53	Cynthia
1/11	2353	847.10	Brett	1/13	2367	91.45	Cynthia
1/4	2302	94.21	Brett	1/6	2324	924.07	Cynthia
1/5	2312	946.08	Brett	1/14	2379	955.23	Cynthia

Darius decided to separate the entries from the check register by employee, and then re-order those entries in increasing order of first digits

Table 4 Expected and actual number of checks written for each first-digit

First Digit	Expected # of checks	# of checks by Brett	# of checks by Cynthia	First Digit	Expected # of checks	# of checks by Brett	# of checks by Cynthia
1	12	12	4	6	3	3	5
2	7	7	5	7	2	2	4
3	5	5	4	8	2	2	4
4	4	4	5	9	2	2	5
5	3	3	4				

It likely would not by itself return a conviction in a court of law, but it does provide support for further investigation of Cynthia's check-writing activities.

In addition to detecting fraud of the type illustrated in this example, Benford's Law may be used to detect instances of cheating on tax returns, filing of false insurance claims, manipulation of a company's stock prices, collusion in bidding processes for government contracts, and a wide variety of other applications.

Benford's Law may be used to detect instances of cheating on tax returns . . . and a wide variety of other applications.

It should also be noted that Benford's Law is more reliably accurate with large sets of data. Some experts recommend it be applied only to data sets of size greater than 100, while others consider even that to be too small. In our fictitious example, intended only to illustrate the general principle, the data set is rather small to be applying Benford's Law with a high degree of confidence.

Using Benford's Law in the classroom

Benford's Law is a topic that students in pre-calculus or above can grasp, and generally find quite interesting. It is a topic that lends itself nicely to individual or group projects in the exploration of real- world data. Students could explore data on a wide variety of topics that might interest them, such as the lengths of rivers of the world, stock prices, census data, or countless other real-world data sets. There are, however, limitations to the type of data sets that Benford's Law applies to. In particular, data that is generated by a formula (for example, student identification numbers that require the first four digits to be the students birth month and date, MMDDxxxxx) would not follow Benford's Law. Likewise, data that is restricted to a small range of numbers (e.g., attendance at a weekly religious service, which is restricted by the building's capacity) would not follow Benford's Law.

In addition to the broad sets of data that can be explored to see if they follow Benford's Law, a conceptual understanding of the law itself can be developed via a compound interest application. We conclude with this example: Imagine someone investing a sum of money, say \$1000, that is being deposited in an account that earns 5% interest per year. After one year, the account will have \$1050; after two years, it will have \$1102.50; and so on. After 14 years, it will have \$1979.95 in it; finally, after year 15, the account will contain \$2078.95. So for 14 years, the lead digit was a 1, because it took that long for the initial investment to double (i.e., increase 100%) and no longer have a lead digit of 1. But beginning in year 15, the investment only has to increase another 50% to move from having a lead digit of 2 to having a lead digit of 3. After an additional seven years, the account has \$2925.30 in it, but the next year it will have \$3071.57. So the lead digit is a 1 for fourteen years, and the lead digit is a 2 for eight more years. Similarly, the lead digit will be a 3 for six years (since it only needs to increase 33% to move from a lead digit of 3 to a lead digit of 4); a 4 for four years,

Table 5 Growth of $1000 compounded at 5% interest per year

Year	Ending Balance ($)	Year	Ending Balance ($)	Year	Ending Balance ($)	Year	Ending Balance ($)
1	1050.00	15	2078.95	23	3071.57	29	4116.21
2	1102.50	16	2182.90	24	3225.15	30	4322.02
3	1157.63	17	2292.05	25	3386.41	31	4538.12
4	1215.51	18	2406.65	26	3555.73	32	4765.03
5	1276.29	19	2526.98	27	3733.52		
6	1340.10	20	2653.33	28	3920.20		
7	1407.11	21	2786.00				
8	1477.47	22	2925.30				
9	1551.34						
10	1628.91						
11	1710.36						
12	1795.88						
13	1885.67						
14	1979.95						

and so on. Students can see Benford's Law develop, in addition to gaining a strong conceptual understanding of it through examples such as this one. Table 5 shows the growth of $1000 compounded at 5% interest per year.

References

Benford, F. (1938). "The Law of Anomalous Numbers." *Proceedings of the American Philosophical Society 78*(4): 551-572.

Newcomb, S. (1881). "Note on the frequency of use of the different digits in natural numbers." *American Journal of Mathematics 4*(1): 39-40.

DARRELL MINOR, dminor@cscc.edu, is Professor of Mathematics at Columbus State Community College. His academic interests include number theory, game theory, the history of mathematics, and providing students with opportunities to discover the beauty of mathematics. Away from work, he enjoys bicycling, reading, and spending time with his wife and three sons.